

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-344553

(43)Date of publication of application : 14.12.2001

(51)Int.Cl. G06F 19/00  
G06F 17/30(21)Application number : 2001-081532 (71)Applicant : SAMSUNG ELECTRONICS  
CO LTD  
REGENTS OF THE UNIV OF  
CALIFORNIA(22)Date of filing : 21.03.2001 (72)Inventor : SHIN HYUN-DOO  
SAI RYORIN  
WU PENG  
MANJUNATH BANGALORE S

(30)Priority

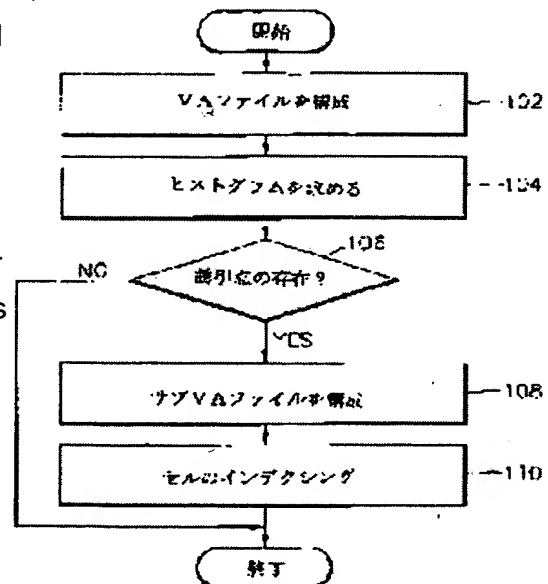
Priority number : 2000 208086 Priority date : 31.05.2000 Priority country : US  
2000 200048323 21.08.2000 KR

## (54) METHOD FOR INDEXING DATA SPACE OF FEATURE VECTOR

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a method for indexing the data space of feature vectors that can be used for similarity retrieval within multi-order vector space.

SOLUTION: This method for indexing the data space of the feature vectors includes a step (a) for discriminating whether a cell where feature vectors are concentrated exists and a step (b) for hierarchically indexing the data spaces of features vectors when the cell where the feature vectors are concentrated exists in the step (a). The method for indexing the data space of the feature vectors can minutely index the data space of the feature vectors when the feature vectors are not equally distributed within the vector space whose order is high.



LEGAL STATUS

[Date of request for examination] 21.03.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-344553

(P 2 0 0 1 - 3 4 4 5 5 3 A)

(43) 公開日 平成13年12月14日 (2001. 12. 14)

(51) Int. Cl. <sup>7</sup>	識別記号	F I	テームコード <sup>*</sup> (参考)
G06F 19/00	140	G06F 19/00	140 5B075
	130		130
17/30	170	17/30	170 G
	220		220 Z

審査請求 有 請求項の数11 O L (全7頁)

(21) 出願番号 特願2001-81532 (P 2001-81532)

(22) 出願日 平成13年3月21日 (2001. 3. 21)

(31) 優先権主張番号 2 0 8 . 0 8 6

(32) 優先日 平成12年5月31日 (2000. 5. 31)

(33) 優先権主張国 米国 (U S)

(31) 優先権主張番号 2 0 0 0 4 8 3 2 3

(32) 優先日 平成12年8月21日 (2000. 8. 21)

(33) 優先権主張国 韓国 (K R)

(71) 出願人 390019839  
三星電子株式会社  
大韓民国京畿道水原市八達区梅灘洞416

(72) 発明者 申 鉉▲ドゥ▼  
大韓民国京畿道城南市盆唐区九美洞221番  
地ムジゲマウル請丘アパート510棟1302号

(72) 発明者 崔 良林  
大韓民国京畿道水原市八達区牛満洞105番  
地牛満鮮京アパート102棟1112号

(74) 代理人 100064908  
弁理士 志賀 正武 (外1名)

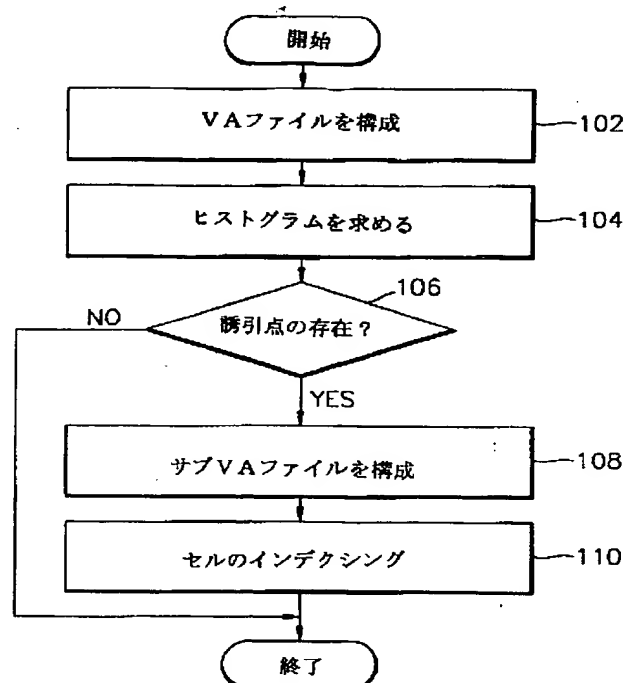
最終頁に続く

(54) 【発明の名称】 特徴ベクトルのデータ空間のインデクシング方法

(57) 【要約】

【課題】 多次ベクトル空間内での類似度の検索に使われる特徴ベクトルのデータ空間のインデクシング方法を提供する。

【解決手段】 特徴ベクトルのデータ空間のインデクシング方法は、(a) 特徴ベクトルが集中したセルが存在しているかを判別する段階及び (b) 前記 (a) 段階において特徴ベクトルが集中したセルが存在すると判別されれば特徴ベクトルのデータ空間を階層的にインデクシングする段階を含むことを特徴とする。特徴ベクトルのデータ空間のインデクシング方法は次数が高いベクトル空間内に特徴ベクトルが均一に分布しない場合に、特徴ベクトルのデータ空間を微細にインデクシングすることが可能である。



## 【特許請求の範囲】

【請求項 1】 特徴ベクトルのデータ空間内にて特徴ベクトルをインデクシングする方法において、

(a) 特徴ベクトルが集中したセルが存在しているかを判別する段階と、

(b) 前記 (a) 段階において特徴ベクトルが集中したセルが存在すると判別されれば、特徴ベクトルのデータ空間を階層的にインデクシングする段階とを含むことを特徴とする特徴ベクトルのデータ空間のインデクシング方法。

【請求項 2】 前記 (a) 段階以前に、

(p a - 1) 前記特徴ベクトルのデータ空間を均等大きさを持つ複数のセルに分割する段階をさらに含むことを特徴とする請求項 1 に記載の特徴ベクトルのデータ空間のインデクシング方法。

【請求項 3】 前記 (a) 段階は、

(a - 1) セル別に特徴ベクトルの数を示すヒストグラムを求める段階と、

(a - 2) 前記ヒストグラムを使用して特徴ベクトルの分布を判別することにより特徴ベクトルが集中したセルが存在しているかを判別する段階とを含むことを特徴とする請求項 1 に記載の特徴ベクトルのデータ空間のインデクシング方法。

【請求項 4】 前記 (b) 段階は、

VA (vector approximation: ベクトル近似化) ファイルを使用してインデクシングする段階を含むことを特徴とする請求項 1 に記載の特徴ベクトルのデータ空間のインデクシング方法。

【請求項 5】 前記 (b) 段階は、

(b - 1) 特徴ベクトルが集中したセルに対しサブ VA ファイルを構成する段階と、

(b - 2) VA ファイルとサブ VA ファイルを使用して該当セル内の特徴ベクトルを近似化する段階とを含むことを特徴とする請求項 4 に記載の特徴ベクトルのデータ空間のインデクシング方法。

【請求項 6】 前記 (b) 段階は、

(b - 1) 前記 (a) 段階において特徴ベクトルが集中したセルが存在すると判別されれば該当セルを分割する段階と、

(b - 1) 前記セル内の特徴ベクトルを分割されたセルを使用して近似化することにより特徴ベクトルのデータ空間を階層的にインデクシングする段階とを含むことを特徴とする請求項 1 に記載の特徴ベクトルのデータ空間のインデクシング方法。

【請求項 7】 特徴ベクトルのデータ空間をインデクシングする方法を行うプログラムコードを保存するコンピュータ読取りの可能な記録媒体において、

(a) 特徴ベクトルが集中したセルが存在しているかを判別する段階と、

(b) 前記 (a) 段階において特徴ベクトルが集中した

セルが存在すると判別されれば特徴ベクトルのデータ空間を階層的にインデクシングする段階とを含む特徴ベクトルのデータ空間のインデクシング方法を行うプログラムコードを保存することを特徴とするコンピュータ読取りの可能な記録媒体。

【請求項 8】 前記 (b) 段階は、

VA ファイルを使用して特徴ベクトルのデータ空間をインデクシングする段階を含むことを特徴とする請求項 7 に記載のコンピュータ読取りの可能な記録媒体。

10 【請求項 9】 前記 (a) 段階は、

(a - 1) セル別に特徴ベクトルの数を示すヒストグラムを求める段階と、

(a - 2) 前記ヒストグラムを使用して特徴ベクトルの分布を判別することにより特徴ベクトルが集中したセルが存在しているかを判別する段階とを含み、

前記 (b) 段階は、

(b - 1) 特徴ベクトルが集中したセルに対しサブ VA ファイルを構成する段階と、

(b - 2) VA ファイルとサブ VA ファイルを使用して該当セル内の特徴ベクトルを近似化する段階とを含むことを特徴とする請求項 8 に記載のコンピュータ読取りの可能な記録媒体。

【請求項 10】 特徴ベクトルがインデクシングされている特徴ベクトルのデータ空間内にて類似度を検索する方法において、

特徴ベクトルが集中したセルが存在しているかを判別し、特徴ベクトルが集中したセルが存在すると判別されれば該当セル内の特徴ベクトルを所定のインデクシング方法により階層的にインデクシングする所定のインデクシング方法によりインデクシングされた特徴ベクトルのデータ空間内にて類似度の検索を行う段階を含むことを特徴とする類似度の検索方法。

【請求項 11】 前記段階は、

NN (nearest neighbor: 最隣接) 検索を基礎とすることを特徴とする請求項 10 に記載の類似度の検索方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は特徴ベクトルのデータ空間のインデクシング方法に係り、より詳細には特徴ベクトルを前記特徴ベクトルの集中度により階層的に近似化することによって、特徴ベクトルの集中度が高いセルに対する細密なインデクシングを可能にする特徴ベクトルのデータ空間のインデクシング方法に関する。

【0002】

【従来の技術】 多量のマルチメディアデータを扱う時には、データベースへの早くて効率的なアクセスが重要である。最近にはマルチメディアデータを生成する能力が早く成長していて、そうしたデータベースを扱ってマルチメディアコンテンツをアクセスする方法が重要な課題

となっている。たとえば、一般的な映像収集は数十万ないし数百万以上のアイテムにわたりなされる。収集された映像を保存するこのようなデータベース内の各客体またはレコードの次数が高い。

【0003】このようなデータベースをアクセスするためには効率的なインデクシング方法を設計することが要求される。インデクシング方法の効率性はそのインデクシング方法に焦点を固定することにより公正に評価される。たとえば、あるインデクシング方法はストリージオーバーヘッドを最小化することに焦点をおく。もう一つのインデクシング方法は照会範囲を効率的に支援することに焦点をおく。

【0004】多次データをインデクシングするためのインデクシング方法が数年間の研究課題になっている。しかし、このような従来のインデクシング方法はマルチメディアデータベースの場合にはドメイン限定的要求により、NN検索を十分に支援するための満足するだけのデータ構造にならなかった。

【0005】このような問題点を解決するための他の従来の技術のインデクシング方法によれば、VAファイルを使用する。しかし、このような従来のインデクシング方法は特徴ベクトルの分布に影響を受けることがある。すなわち、前記のような従来のインデクシング方法によれば、特徴ベクトルが均一に分布する場合には顕著に簡便化することが可能であるが、特徴ベクトルが均一に分布しない場合にはインデクシングが効率的になされない場合があるという問題点がある。

#### 【0006】

【発明が解決しようとする課題】本発明がなそうとする技術的課題は特徴ベクトルの集中度が高いセルに対する細密なインデクシングを可能にする特徴ベクトルのデータ空間のインデクシング方法を提供することである。本発明がなそうとする他の技術的課題は前記特徴ベクトルのデータ空間のインデクシング方法を行うプログラムコードを保存するコンピュータ読取りの可能な記録媒体を提供することである。本発明がなそうとするさらに他の技術的課題は前記特徴ベクトルのデータ空間のインデクシング方法によりインデクシングされた特徴ベクトルのデータ空間内にて類似度の検索を行う類似度の検索方法を提供することである。

#### 【0007】

【課題を解決するための手段】前記課題をなすために本発明による特徴ベクトルのデータ空間のインデクシング方法は、(a) 特徴ベクトルが集中したセルが存在しているかを判別する段階及び(b) 前記(a) 段階において特徴ベクトルが集中したセルが存在すると判別されれば特徴ベクトルのデータ空間を階層的にインデクシングする段階を含むことを特徴とする。

【0008】さらに、前記特徴ベクトルのデータ空間のインデクシング方法は、前記(a) 段階以前に、(p a

ー1) 前記特徴ベクトルのデータ空間を均等な大きさを持つ複数のセルに分割する段階をさらに含むことが望ましい。さらに、前記(a) 段階は、(a-1) セル別に特徴ベクトルの数を示すヒストグラムを求める段階及び(a-2) 前記ヒストグラムを使用して特徴ベクトルの分布を判別することにより特徴ベクトルが集中したセルが存在しているかを判別する段階を含むことが望ましい。

【0009】さらに、前記(b) 段階は、VAファイルを使用してインデクシングする段階を含むことが望ましい。さらに、前記(b) 段階は、(b-1) 特徴ベクトルが集中したセルに対しサブVAファイルを構成する段階と、(b-2) VAファイルとサブVAファイルを使用して該当セル内の特徴ベクトルを近似化する段階とを含むことが望ましい。

【0010】さらに、前記(b) 段階は、(b-1) 前記(a) 段階において特徴ベクトルが集中したセルが存在すると判別されれば該当セルを分割する段階及び(b-1) 前記セル内の特徴ベクトルを分割されたセルを使用して近似化することにより特徴ベクトルのデータ空間を階層的にインデクシングする段階を含むことが望ましい。さらに、前記他の課題を成し遂げるために本発明によるコンピュータ読取りの可能な記録媒体は、(a) 特徴ベクトルが集中したセルが存在しているかを判別する段階及び(b) 前記(a) 段階において特徴ベクトルが集中したセルが存在すると判別されれば特徴ベクトルのデータ空間を階層的にインデクシングする段階を含む特徴ベクトルのデータ空間のインデクシング方法を行うプログラムコードを保存することを特徴とする。

【0011】さらに、前記さらに他の課題をなすために本発明による類似度の検索方法は、特徴ベクトルが集中したセルが存在しているかを判別し、特徴ベクトルが集中したセルが存在すると判別されれば該当セル内の特徴ベクトルを所定のインデクシング方法により階層的にインデクシングする所定のインデクシング方法によって、インデクシングされた特徴ベクトルのデータ空間内にて類似度の検索を行う段階を含むことを特徴とする。

#### 【0012】

【発明の実施の形態】以下、添付された図面を参照して本発明の望ましい実施形態を詳細に説明する。図1には本発明の実施形態によるインデクシング方法の主要段階をフローチャートに示した。図1を参照すれば、本発明の実施形態によるインデクシング方法では、まず全体特徴ベクトルのデータ空間内にてVAファイルを構成する(段階102)。VAファイルを構成するために、前記特徴ベクトルのデータ空間は均等な大きさを持つ複数のセルに分割されている。本明細書では本発明が効果的に作用する状況を説明するために特徴ベクトルが分割された複数のセルのうちから任意のセルに集中している場合を仮定する。

【0013】図2にはVAファイルを構成するための特徴ベクトルのデータ空間の一例を示した。図2を参照すれば、内部の特徴ベクトルが、0101に近似化されるセル20と1011に近似化されるセル22に特徴ベクトルが集中している。以下ではこのように特徴ベクトルが集中しているセルを誘引点と称する。

【0014】次に、全体ベクトル空間に対し特徴ベクトルの分布を示すヒストグラムを求める（段階104）。さて、前記ヒストグラムを参照して誘引点が存在しているかを決定する（段階106）。たとえば、ヒストグラム上に特徴ベクトルの数が所定数以上のセルを誘引点として決定することが可能である。本実施形態では10以上のセルを誘引点として決定する場合を例として説明する。すなわち、内部の特徴ベクトルが0101に近似化されるセル20と1011に近似化されるセル22内で特徴ベクトルの数が10以上であると示されるので、該当セルを誘引点に該当するセルとして決定する。

【0015】次に、誘引点が存在すれば該当セルに対しサブVAファイルを構成する（段階108）。該当セルは複数のサブセルに分割される。サブVAファイルはサブセル内に存在する特徴ベクトルの位置を参照して決定される。図3(a)と図3(b)には誘引点が存在するセルを複数のサブセルに分割した例を説明するための図面を示した。図3(a)と図3(b)をそれぞれ参照すれば、0101に近似化されたセルと1011に近似化されたセルは複数のサブセルに分割されている。サブセル内に存在する特徴ベクトルの位置を参照してサブVAファイルが構成される。

【0016】一方、誘引点が存在しなければベクトル空間内に近似的に均一性が維持されていると言えることを意味するので、一般的なVAファイルが使われる。すなわち、特徴ベクトルのデータ空間内の特徴ベクトルを分割されたセル別にそのまま近似化されてVAファイルを構成する。さて、VAファイルとサブVAファイルを使用して該当セル内の特徴ベクトルを近似化する。たとえば、0101として参照されるセル内に存在する特徴ベクトルデータ302と特徴ベクトルデータ304は、それぞれ01010110と01010111に近似化される。さらに、1011として参照されるセル内に存在する特徴ベクトルデータ322と特徴ベクトルデータ324は、それぞれ10110001と10111010に近似化される。こうして、該当セルのサブVAファイルが構成されることにより、該当セルはVAファイルとサブVAファイルとが結びついたファイルとしてインデクシングされる。VAファイルとサブVAファイルとが結びついたファイルはHVAファイルと言える。

【0017】前記のようなインデクシング方法によれば、特徴ベクトルのデータ空間を特徴ベクトルの分布を参照して階層的に近似化することにより該当セルをイン

デクシングする。階層的なインデクシングは特徴ベクトルの集中度が高いセルに対する細密なインデクシングを可能にする。特に、前記のような本発明による方法によれば、次数が高いベクトル空間内に特徴ベクトルが均一に分布しない場合に、前記特徴ベクトルをより効率的にインデクシングする。言い換えれば、特徴ベクトルデータが集中する場合に対処するために特徴ベクトルのデータ空間内の特徴ベクトルデータの分布により近似化構造が調整される。

【0018】以下では図1を参照して説明した本発明による特徴ベクトル空間のインデクシング方法により階層的にインデクシングされた特徴ベクトルのデータ空間内にて類似度の検索を行う過程を説明する。本発明の実施形態による類似度の検索方法によれば、図1を参照して説明した本発明による特徴ベクトルのデータ空間のインデクシング方法を使用してインデクシングされた特徴ベクトルのデータ空間内にて類似度の検索を行う。前記特徴ベクトルのデータ空間内にて特徴ベクトルが集中しているセル内の特徴ベクトルは、サブVAファイルを使用して近似化されている。たとえば、01010110に近似化された照会点に対し類似度の検索を行う場合には、特徴ベクトルのデータ空間内にて0101に近似化されたセルが検索されたセルとして選択され、前記セル内で0110に近似化されたセルが存在しているかを判別する。もし、前記セル内に0110に近似化されたセルが存在すると判別されれば該当セルが検索されたセルとして決定される。

【0019】前記のような類似度の検索方法は次数が高いベクトル空間内に特徴ベクトルが均一に分布しない場合にも特徴ベクトルのデータ空間内にて照会点と類似した特徴を持つ特徴点を細密で正確に検索することが可能である。検索方法にはNN検索を含んだ多様な検索方法が適用されうる。図1を参照して説明した実施形態では、二段階の階層的インデクシングを行うことを例として説明したが、より多くの段階の階層的インデクシングを行うことが可能である。さらに、図1を参照して説明した実施形態では、誘引点が存在しているかを判別するためにヒストグラムを使用することを例として説明したが、当業者により適切な他の分析方法を使用するように変更または修正することが可能である。すなわち、請求項により限定される本願発明の範囲は前記実施形態に限定されない。

【0020】さらに、前記のような本発明によるインデクシング方法は個人用またはサーバ級のコンピュータ内で実行されるプログラムで作成可能である。前記プログラムを構成するプログラムコード及びコードセグメントは当該分野のコンピュータプログラマにより容易に推論できる。さらに、前記プログラムはコンピュータ読取りの可能な記録媒体に保存されうる。前記記録媒体は磁気記録媒体、光記録媒体及び電波媒体を含む。

## 【0021】

【発明の効果】 前述したように本発明による特徴ベクトルのデータ空間のインデクシング方法は、次数の高いベクトル空間内に特徴ベクトルが均一に分布しない場合に特徴ベクトルのデータ空間を微細にインデクシングすることが可能である。さらに、本発明による類似度の検索方法は、次数の高いベクトル空間内に特徴ベクトルが均一に分布しない場合にも特徴ベクトルのデータ空間内で照会点と類似した特徴を持つ特徴点を細密で正確に検

索することが可能である。

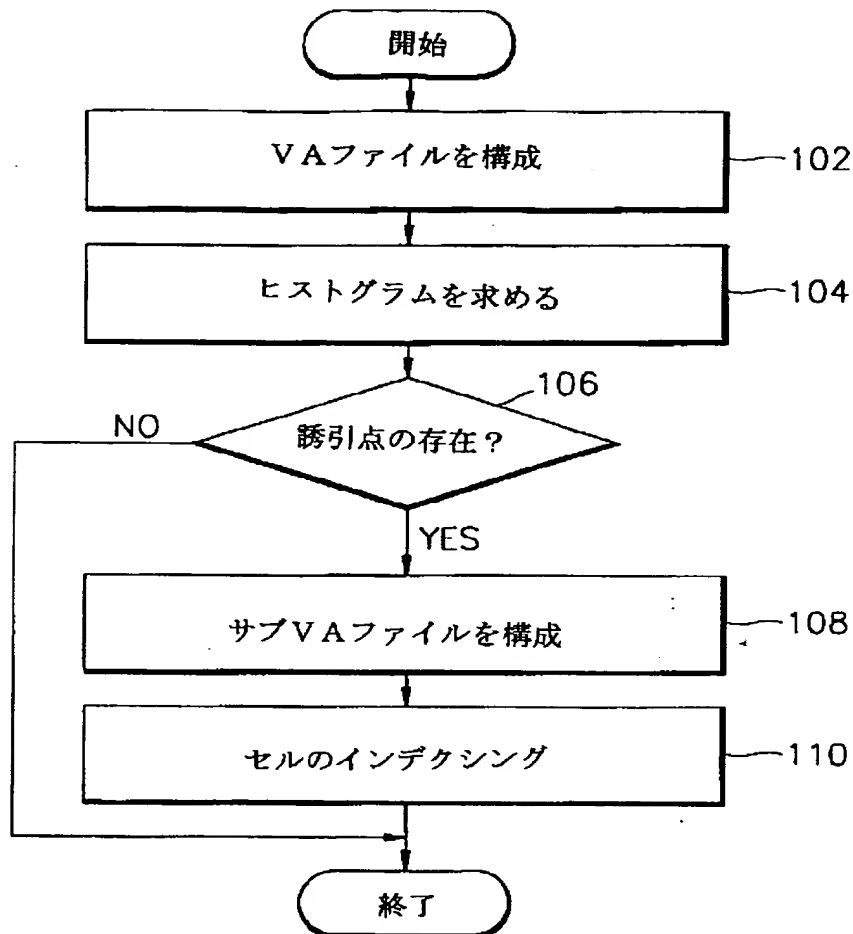
## 【図面の簡単な説明】

【図1】 本発明の実施形態による特徴ベクトルのデータ空間のインデクシング方法の主要段階を示したフローチャートである。

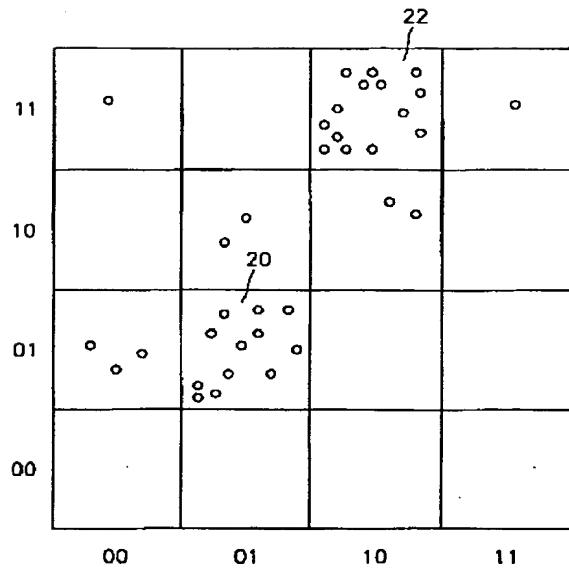
【図2】 VAファイルを構成するための特徴ベクトルのデータ空間の一例を示した図面である。

【図3】 誘引点が存在するセルを複数のサブセルに分割した例を説明するための図面である。

【図1】

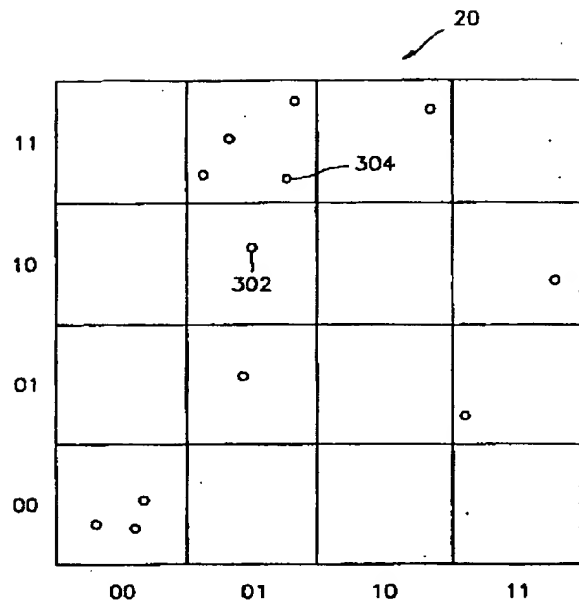


【図 2】

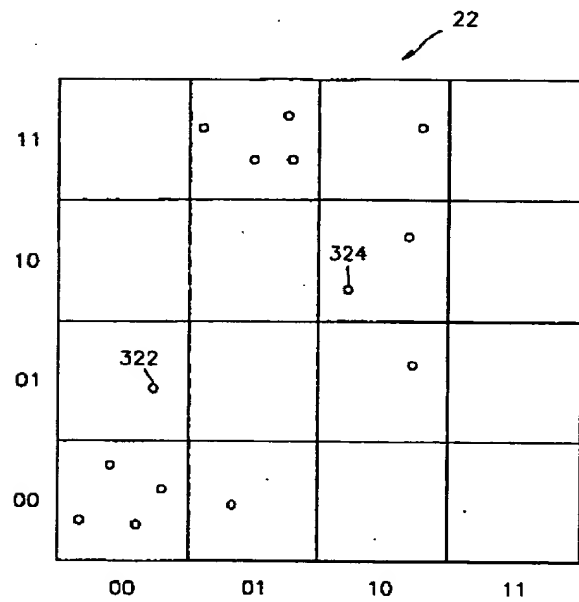


【図 3】

(a)



(b)





フロントページの続き

(71)出願人 501276728

ザ・リージェンツ・オブ・ザ・ユニバーシ  
ティー・オブ・カリフォルニア  
THE REGENTS OF THE  
UNIVERSITY OF CALIF  
ORNIA

アメリカ合衆国、カリフォルニア州94607  
-5200、オークランド、フランクリン・ス  
トリート1111、トゥエルブス・フロア

(72)発明者 吳 澎

アメリカ合衆国・93106-9560・カリフォ  
ルニア・サンタバーバラ・ユニヴァーシテ  
ィー・オブ・カリフォルニア・ディパート  
メント・オブ・エレクトリカル・アンド・  
コンピューター・エンジニアリング

(72)発明者 バンガロア・エス・マンジュナス

アメリカ合衆国・93106-9560・カリフォ  
ルニア・サンタバーバラ・ユニヴァーシテ  
ィー・オブ・カリフォルニア・ディパート  
メント・オブ・エレクトリカル・アンド・  
コンピューター・エンジニアリング

Fターム(参考) 5B075 ND16 QS20